ERDC/CHL CHETN-I-86
July 2015

US Army Corps of Engineers®

Computer-Based Model Calibration and Uncertainty Analysis: Terms and Concepts

by Alejandro Sánchez, Brian Skahill, William Scharffenberg, and Cole H. Smith

PURPOSE: The purpose of this document is to provide a description of key terms and concepts specific to computer-based model calibration and uncertainty analysis. This document is designed for engineers, scientists, and planners who perform or utilize environmental modeling for planning, restoration, and operations purposes. More specifically, the document is intended to facilitate the understanding and promote the proper use of model calibration and uncertainty analysis within the U.S. Army Corps of Engineers (USACE), Hydrologic and Hydraulic (H&H) Community of Practice (CoP).

PREFACE: The USACE is required to perform risk and uncertainty analyses throughout the lifecycle of planning, designing, and operating of Civil Works flood risk management projects as described in USACE (1996a,b,c). The reference documents describe how to quantify uncertainty and incorporate it into economic and engineering performance analysis. The purpose of the USACE risk analysis policy is to improve decision making and create confidence in the project formulation and evaluation process by quantifying risk and disclosing uncertainty in key data and parameters (Davis et al. 2008). However, the current capacity within the USACE to quantify the uncertainty of H&H model predictions, and consequently accurately compute risk, is limited (Skahill 2013a). Recent efforts to improve the USACE use of risk-based analysis in H&H studies have produced several publications and tools (HEC 2006; Srivastava 2008; Skahill and Baggett 2012; Skahill 2013a,b). Scharffenberg and Kavvas (2011) estimated the uncertainty in flood wave routing using a stochastic model with external input from the Hydrologic Engineer Center (HEC) Hydrologic Modeling System (HEC-HMS). Skahill (2013a) described and reviewed several state-of-the-art and practice-oriented uncertainty analysis methods, with the purpose of identifying the most promising approaches for H&H applications. Skahill (2013b) provided a path forward for related work activities, including software development, preparation of practice-oriented guidance documentation, and research and development directed at improving uncertainty analysis algorithm efficiency. This document provides a brief overview of basic concepts, terms, and techniques used in computer-based model parameter calibration and uncertainty quantification. References are provided with further details for the interested reader.

INTRODUCTION: Numerical models and their simulation are powerful and useful tools because they provide a way of understanding the underlying mechanisms that control system behavior. Moreover, from a practical perspective, they provide a cost-effective tool to conduct predictions (forecasts) and evaluate alternative designs, plans, or policies. Invariably, all H&H models, either empirical or physics-based, have variable input parameters which need to be estimated. In the best of cases, parameters can be estimated directly from measurements. However, models often contain parameters which cannot be measured directly either because they have no physical basis, it would be expensive, or they are difficult to measure due to spatial

Report Documentation Page

Form Approved OMB No. 0704-0188

1. REPORT DATE **JUL 2015**	2. REPORT TYPE	3. DATES COVERED **00-00-2015 to 00-00-2015**

4. TITLE AND SUBTITLE **Computer-Based Model Calibration and Uncertainty Analysis: Terms and Concepts**	5a. CONTRACT NUMBER 5b. GRANT NUMBER 5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER 5e. TASK NUMBER 5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **U.S. Army Engineer Research and Development Center,,Vicksburg,,MS**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR’S ACRONYM(S) 11. SPONSOR/MONITOR’S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The purpose of this document is to provide a description of key terms and concepts specific to computer-based model calibration and uncertainty analysis. This document is designed for engineers, scientists, and planners who perform or utilize environmental modeling for planning, restoration, and operations purposes. More specifically, the document is intended to facilitate the understanding and promote the proper use of model calibration and uncertainty analysis within the U.S. Army Corps of Engineers (USACE), Hydrologic and Hydraulic (H&H) Community of Practice (CoP).

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**12**	

averaging, among other reasons. Observations of system state (e.g., stages and flows) are typically more cost effective to obtain and/or readily available than measurements of model input process related parameters, assuming that in fact they are measurable. Model calibration is the process of inversely determining uncertain model input parameters by varying the parameter values themselves in attempts to improve the match between the measured indirect variables and their simulated counterparts. The intended purpose(s) of a given model deployment and data quality are primary factors that influence potentially predetermined criteria indicating a model calibration to be complete. Hence, a model deployment may employ several independent calibrations depending on the specific model purpose or use. It is noted that not all calibrations result in an acceptable level of model uncertainty; in which case it may be necessary to collect more measurements or improve the model.

Model calibration performed by manually testing parameter values by trial-and-error is commonly employed in the H&H CoP. However, manual calibration is subjective, labor intensive, and lacks repeatability. In contrast, computer-based model calibration, or for succinctness, parameter estimation (PE), is more objective, repeatable, and better capitalizes on the computational capacity of the modern computer. One of the most important benefits of computer-based model calibration is that its application results in model input uncertainty and model predictive uncertainty that can be used to support risk-informed decision making. Moreover, when a model contains a large number of estimable parameters, manual calibration can become technically infeasible. For example, modern hydrologic model deployments can easily have a large number of calibration parameters due to empirical process formulations that are parameterized based on high-resolution spatial data products. Computer-based calibration methodologies often provide a means to effectively calibrate highly parameterized models. However, care must be exercised when employing computer-based methodologies to ensure a quality modeling product with useful outputs.

H&H models are only approximate representations of the physical environment and have several sources of error and uncertainty including

- inaccuracy or indeterminacy of input parameters (experimental or field measurements, environment variability, incompatibility of scales)
- lack of knowledge of processes (e.g., future watershed management practices)
- approximations in the mathematical model which may neglect certain processes (e.g., soil models generally do not include individual tree roots or animal burrows that may affect infiltration in certain situations)
- uncertainty in describing the physical reality (e.g., errors in geometry, boundary conditions).

Uncertainty analysis (UA), as defined here, is the formal process of characterizing model input uncertainty and mapping it onto model output uncertainty, hence, quantifying the range of likely outcomes. In other words, UA determines how likely specific outcomes are if some aspects of the system are not exactly known. UA supports the following activities:

- understanding of the predictive capabilities and limitations of a model
- risk-informed decision making

- informing the value of collecting additional measurements with the intent of reducing model uncertainty
- providing a basis for extrapolating model parameters to similar problems/conditions for which measurements are not available
- providing a potential basis for model comparison and selection.

Not surprisingly, formal methods have been developed for PE and UA. Reviews of PE and UA concepts and methods may be found in Skahill (2013a), Hill (1998), Mishra (2009), Matott et al. (2009), and Vanrolleghem (2010). There are several commercial and open-source software packages available for PE and UA (e.g., Beven and Binley 1992; Doherty 2004; Skahill et al. 2009). As mentioned previously, Skahill (2013a) described and reviewed several state-of-the-art and practice-oriented uncertainty analysis methods. Matott et al. (2009) provide a comprehensive review of related software-based tools.

Typically, an environmental modeling study will consists of the following steps (Vanrolleghem 2010):

1. Plan development
2. Data collection and conceptualization
3. Model set-up
4. Calibration and validation
5. Simulation and evaluation.

This document is aimed at describing key terms and concepts related to the methods and techniques applied in computer-based model calibration, and estimation of parameter and predictive uncertainty as part of steps 4 and 5 listed directly above.

TERMS AND CONCEPTS: The following subsections describe key terms and concepts related to computer-based model calibration and uncertainty analysis.

Parameter estimation. Parameter estimation (PE), or model calibration, is the process of inversely determining uncertain model input parameters by varying the parameter values themselves in attempts to improve the match between a set of historical observations of the system being modeled and their simulated counterparts. An example is adjusting model parameters to reduce the difference between computed and observed stream flow. Conceptually, a calibration parameter can include any uncertain value that is used by a model, irrespective of whether the quantity to which this value is assigned is a system property, a boundary condition, or an aspect of the model's geometry. PE methods may be classified as single-solution or multiple-solution (Matott et al. 2009). PE generally requires the specification of adjustable parameters, calibration data, and objective function(s).

Objective function. An objective function is a user-defined quantitative performance measure depicting how well observations of system state(s) compare with their model simulated counterparts (Servat and Dezetier 1991; Skahill 2006). Generally, PE methods work by minimizing or maximizing the specified objective function. The objective function may also be constructed from multiple variables or the same variable processed in different ways (Madsen 2000; Boyle et al. 2000; Doherty and Johnston 2003). In some cases, the data is transformed before fitting to

compensate for different magnitudes in variables (Kuczera 1983; Bates and Campbell 2001). Objective function formulation is both science and art, and moreover, during a model calibration, it may take several iterations before a satisfactory final definition for the objective function is identified. One possible example, among many, is the root-mean-square error.

Single-solution PE methods. Single-solution PE methods yield point estimates of parameter values and do not directly provide estimates of parameter uncertainty. Single-solution PE methods may be grouped as local, global, and hybrid search methods. Local search methods are optimization algorithms which progressively adjust initial parameter estimates towards an optimal solution (hopefully). Local search methods may converge in a locally optimum value depending on the initial solution. Finding the globally optimum value is more difficult but may be accomplished using global search methods. Hybrid search methods utilize several concurrent local search methods with different initial solutions in an attempt to ensure finding the global optimal solution. Each approach's strengths and weaknesses, together with a given project's objectives and available resources, need to be taken into consideration when choosing a specific single-solution PE method for model calibration. In order to estimate parameter confidence intervals, single-solution methods may be combined with postcalibration analysis.

Multiple-solution PE methods. Application of a multiple-solution PE method simultaneously results in, as its name suggests, a set of possible solutions and estimates of model uncertainty. Multiple-solution PE methods are Monte Carlo-based approaches and may be categorized as (1) importance sampling and (2) Markov chain Monte Carlo (MCMC) sampling. Multiple-solution PE methods are generally more computationally intensive than single-solution methods but have the advantage of providing parameter and predictive uncertainty estimates.

Sensitivity analysis. A sensitivity analysis (SA) computes the effect on the simulated outputs resulting from specified changes to model input values. Several SA methods exist; however, the most common approach is to determine the change in model response(s) given a perturbation from a base value for each individual model parameter. The computed model sensitivity may then be scaled by the range of the individual parameter or the model response in order to compare different parameter sensitivities. The model response may be a specific model output variable or aggregate measure. An SA is often employed to select and/or reduce the number of adjustable model parameters to be considered as adjustable for model calibration. However, it is possible that individual parameter sensitivities do not determine what is estimable and what is not. Situations are often encountered where model outputs have a low sensitivity to certain parameters collectively but can be very sensitive to the same parameters individually. This is the phenomenon of *parameter correlation*. Moreover, it is possible for parameter(s) to be insensitive for a calibration dataset but sensitive for a validation dataset, indicating the calibration dataset is not adequate. This will lead to less-than-optimal predictive capabilities.

Parameter correlation. Parameter correlation refers to the statistical dependence between two parameters. Parameter correlation is expressed as the off-diagonal terms of the parameter covariance matrix. Positive values indicate positive correlation (i.e., a higher value in one parameter will produce a higher value in the other), and negative values indicate negative correlation (i.e., a lower value in one parameter will produce a higher value in the other). Parameter correlation is an important aspect of PE because it can lead to instability or slow

convergence rates. Parameter correlation also invalidates the use of common error propagation equations of Mandel (1984).

Principle of parsimony. The principle of parsimony states that the parameters to specify as adjustable during calibration should be limited to those for which a unique estimate can be obtained. One approach for adhering to the principle of parsimony is to start with only a few adjustable model parameters while maintaining others at fixed values. Then, estimable parameters are added sequentially to the extent that they result in an improved fit to the observed data, and the calibration diagnostic information does not indicate increased model (i.e., parameter) uncertainty and/or numerical instability of the parameter estimation process itself. The selection of model parameters can be guided by a previous calibration study or a sensitivity analysis. An alternate approach is to start with a highly parameterized model and enforce the principle of parsimony in the calibration process by explicitly including a related penalty term in the objective function (e.g., Doherty 2004). This approach stabilizes what may otherwise be an unstable model calibration and permits the modeler(s) to explore the tradeoff between enforcing adherence to a preferred parsimonious model state with the fit to the observed data.

The principle of parsimony is important; however, in practice it can be difficult to adhere to with modern numerical model deployments that simulate numerous physical processes with parameters whose values are estimated based on input spatial data products. Strict adherence to the principle of parsimony can result in poor fits to the observed data and, moreover, a final model that is not useful as an analysis tool to address its originally intended objectives. Modern environmental model deployments must often accommodate complex requirements that necessitate the dimensionality of the estimable model parameter space to not necessarily be in line with the principle of parsimony. Conversely, selecting too many calibration parameters may produce good calibration results but obtained with a final estimated model that is deemed nonphysical. At the extreme, a highly parameterized model may result in an ill-posed inverse model calibration problem wherein there are more model parameters than there are observations to support their unique estimation. Parameters that cannot be uniquely estimated can be assigned values or relationships with other parameters (Doherty and Skahill 2006). These cases require special calibration methods as compared to those typically used for well-posed or over-determined models (models with a unique and well-defined calibration parameter set).

Regularization. Regularization is a model inversion technique which can be used to ensure that a stable solution is obtained to an otherwise ill-posed inverse problem. In practice, regularization is achieved through adherence to the principle of parsimony. Regularization may also be applied by lumping, grouping, or averaging environmental processes. This is referred to as structural regularization.

Model equifinality. This is a concept in which there are multiple equally acceptable models given possibly ambiguous data and the incomplete understanding of a system(s). The models may differ in their basic structure or only in their calibration parameters. Models are equifinal if they lead to equally acceptable representations of a system(s) even though their structure or parameters are different.

Heteroscedasticity. This term means "differing variance" indicating that a collection of random variables has subpopulations each with different variance. Thus, heteroscedasticity is the

absence of homoscedasticity. The possible existence of heteroscedasticity is a major concern in the application of regression analysis, including the analysis of variance, because the presence of heteroscedasticity can invalidate statistical tests of significance that assume that the modelling errors are uncorrelated and normally distributed and that their variances do not change with the effects being modeled. Similarly, in testing for differences between subpopulations using a location test, some standard tests assume that variances within groups are equal.

Serial correlation. Serial correlation occurs when error terms from different (usually adjacent) temporal or spatial data are correlated. For example, in predicting a river stage, an overestimate at one time or location will generally lead to an overestimate at a subsequent time or downstream location, respectively. In comparing model results to measurements, serial correlation between adjacent time-steps (short-range autocorrelation), can be handled by comparing the rate of change (or slope) of the observed and modeled values.

Model parameterization. Model parameterization generally refers to the process of determining the functions or methods to represent the spatial and/or temporal variation of model variables. In model calibration, parameterization is a useful technique to reduce the number of adjustable parameters and *parsimonize* the problem. A review of various parameterization methodologies is given in Vanrolleghem (2010).

Uncertainty analysis. Uncertainty analysis (UA) generally includes both uncertainty characterization and propagation. Uncertainties relate to the physics of the problem of interest not to the errors in the mathematical discretization and numerical solution techniques. UA methods may be classified as frequentist or Bayesian. The types of uncertainty are summarized below. UA allows estimating distributions around model input and output.

Uncertainty characterization. Uncertainty characterization is the process of fitting and/or assigning marginal and joint distributions to uncertain model inputs.

Uncertainty propagation. Uncertainty propagation is the process of translating the uncertainty in model inputs into the corresponding uncertainty in model outputs.

Epistemic uncertainty. Epistemic uncertainty is a type of uncertainty which can be reduced by increasing the knowledge or understanding of the system and improving its mathematical representation (Ross et al. 2009).

Stochastic uncertainty. Stochastic uncertainty, also referred to as aleatory uncertainty or first-order uncertainty, arises from natural variability of the system (Ross et al. 2009; Briggs et al. 2012). Hence, it cannot be reduced, and additional experiments or measurements can only be used to better characterize the variability. Stochastic uncertainty is analogous to the error term in regression analysis.

Parameter uncertainty. Parameter uncertainty, also referred to as second-order uncertainty, refers to the input parameter and forcing (boundary conditions, etc.) uncertainty (Vanrolleghem 2010; Briggs et al. 2012). Parameter uncertainty is analogous to the standard error of the estimate in regression analysis. Parameter uncertainty is a type of epistemic uncertainty.

Structural uncertainty. Structural uncertainty, also referred to as model uncertainty, refers to the uncertainty related to the inherent assumptions in the formulation of the mathematical model itself (i.e., the model does not perfectly represent the system) (Briggs et al. 2012). Structural uncertainty is a type of epistemic uncertainty and is analogous to the form of the regression model in regression analysis.

Model error. Model errors are related to the solution techniques of a mathematical model such as discretization and truncation error, round-off, numerical instabilities.

Linear UA. Linear (first-order) UA is a postcalibration analysis that may be used in combination with a single-solution PE method to obtain relative parameter uncertainty estimates, parameter correlation coefficients, parameter sensitivities, and limited predictive uncertainty estimates. As stated by Gallagher and Doherty (2007), a linear UA may be calculated with minimal additional computational costs to the PE.

Nonlinear UA. Nonlinear (second-order) UA is a postcalibration analysis that may be used in combination with a single-solution PE method to obtain parameter and predictive uncertainty estimates, parameter correlation coefficients, and parameter sensitivities. Examples of nonlinear UA methods are the calibration-constrained optimization analysis (Tonkin et al. 2007) and the calibration-constrained Monte Carlo analysis (Tonkin et al. 2009).

Importance sampling. Importance sampling is a multiple-solution PE method in which sampled model parameter sets are divided into behavioral and nonbehavioral groups, based on an acceptance threshold for an objective function. The nonbehavioral group is discarded or rejected, and model parameter distributions are estimated using a weighted or bias-corrected combination of the behavioral parameter group. An example of an importance sampling PE and UA method is the Generalized Likelihood Uncertainty Engine (GLUE) (Beven and Binley 1992). Additional information on GLUE may be found in Skahill (2013a) and Vrugt et al. (2008).

Frequentist PE and UA. Frequentist probability is objective and refers to an event's relative frequency in a large number of trials. For example, a coin has a 50% probability of falling heads. Likewise, the 95% confidence interval of a parameter will contain the true parameter value 95% of the time. In the frequentist approach to PE, model parameters area regarded as having true values, and their estimate is based on the calibration dataset. Intervals are calculated for the parameters indicating the level of confidence (e.g., 95%) of the true value falling within that interval. In the frequentist approach to UA, the probability distributions of input parameters are assigned and held constant. The parameter distributions are then sampled, and the model results are assembled to produce probability distributions of model results or outcomes. As mentioned previously, this process of moving from input uncertainty to uncertainty in output is referred to as uncertainty propagation.

Monte Carlo Simulation. The Monte Carlo Simulation (MCS) or Monte Carlo Method is in actuality a broad class of computational algorithms which employ random sampling algorithms used to obtain probability distributions of an unknown variable (Robert and Casella 2004). MCS is a commonly used method for uncertainty propagation. In MCS, the uncertain parameters are sampled from distributions, and a large number of simulations are computed. Each simulation represents an independent and equally probable realization. MCS has several benefits: (1) it

avoids errors associated with linearization of the model; (2) it produces a distribution for the uncertain output as well as the mean and standard deviation; (3) the method does not require the computation of derivatives; and (4) can handle correlated and independent parameters. However, MCS may not be the most efficient in terms of number of model runs when the parameter uncertainty is poorly defined or the outcomes of interest are limited in number (Mishra 2003).

Bayesian PE and UA. In Bayesian PE and UA, probability distributions are used to quantify the uncertainty in model parameters. Initially, prior probability distributions are assigned that depend on any initial information available. The prior probability distributions are updated using Bayes' theorem to produce posterior probability distributions for the parameters based on new data or evidence. The probability estimate of a hypothesis is expressed in terms of degrees of belief or more specifically, viz. Bayesian probabilities. Bayesian inference, the process of making conclusions from data that are subject to random variation, is based on Bayes' rule, which is used to update the probability estimate for a hypothesis as new evidence is obtained.

A significant advantage of Bayesian PE and UA is that it allows for the incorporation of both hard data (i.e., quantitative) and soft data (i.e., qualitative data) into an analysis. The Bayesian approach evaluates the probability of a hypothesis by specifying some prior probability, which is then updated with new evidence (Bayes' rule) (Lee 2012). Bayesian probability is subjective and can be applied to single events based on the degree of confidence or belief. For example, in a Bayesian framework, someone could say that tomorrow's weather has a 50% chance of rain. Whereas in a frequentist framework, someone can only say that there is a certain probably of rain for a given day of the year based on the historical record. Parameters in a Bayesian simulation are random variables that have a prior and posterior distribution.

Bayes' rule. Bayes' rule is given by $p(\theta|y) = L(y|\theta)p(\theta)$ where p() indicates probability, $p(\theta|y)$ is the posterior probability distribution of the parameters θ, $L(y|\theta)$ is the likelihood function, and $p(\theta)$ is the prior probability density function (PDF). The prior PDF, $p(\theta)$, represents information about θ before any data are collected. $L(y|\theta)$ encapsulates the new data or evidence.

Likelihood function. In Bayesian statistics, the likelihood function encapsulates the new data or evidence. The likelihood is the distribution of the observed data conditional on its parameters.

Markov Chain Monte Carlo. Markov Chain Monte Carlo (MCMC) is a class of statistical methods used for sampling from probability distributions based on constructing Markov chains. A Markov chain is a stochastic process of values that unfold in time with an equilibrium distribution equal to the desired probability density. Therefore, a Markov chain can be constructed and run until equilibrium, and the probability density function can be obtained by sampling from its stationary distribution. A Markov chain can be constructed by choosing a symmetric proposal distribution and employing the Metropolis acceptance probability (Metropolis et al. 1953) to accept or reject candidate points. Unlike in the traditional Monte Carlo method, where the random samples are statistically independent, the samples in MCMC are correlated. MCMC is generally more efficient than other Monte Carlo methods. The ability to sample from the posterior probability distribution for the specified adjustable model parameters provides the capacity to robustly address questions associated with the deployed modeled scenarios/alternatives from a probabilistic perspective.

Burn-in. In MCMC, burn-in is the colloquial term that describes the practice of throwing away a certain number of iterations at the beginning of an MCMC run. The intent of the burn-in is to remove the effect of the initial sample (starting point), which may have a low probability in the equilibrium distribution. An alternative to the burn-in is to start a Markov chain using the last iteration of a previous MCMC run.

Verification. This is the process of determining that a model implementation accurately represents the intended conceptual description of the model (i.e., are the governing equations formulated and solved correctly). Verification may be done using analytical solutions to the governing equations for idealized conditions.

Validation. This is the process of determining the degree to which a model is an accurate representation of the real world for the intended uses of the model (i.e., are the correct equations being solved). Once a model is calibrated using a set of observations or measurements, it is recommended to validate the model using a different set of measurements. It is important that the validation data be different from the calibration data. A common approach is to use the same type of data from the same location(s) used during calibration but for a different time window. An alternative is to use the same time window but to use different data types and/or observation locations that were not used during calibration.

CONCLUSIONS: Key terms and concepts related to computer-based model calibration and uncertainty analysis have been described and references provided with further details. The document is intended to familiarize and promote the proper use of PE and UA methods within the USACE H&H CoP. Subsequent work will include the development of guidelines for the use of PE and UA software in combination with H&H models.

ADDITIONAL INFORMATION: Questions about this CHETN can be addressed to Alejandro Sánchez at (601-634-2027), FAX (601-634-3433), or e-mail: *Alejandro.Sanchez@usace.army.mil*. An electronic copy of this CHETN is available from *http://chl.erdc.usace.army.mil/chetn/*. This Technical Note should be referenced as follows:

> Sánchez, A., B. E. Skahill, W. Scharffenberg, and C. H. Smith. 2015. *Computer-based model calibration and uncertainty analysis: Terms and concepts*. ERDC/CHL CHETN I-86. Vicksburg, MS: U.S. Army Engineer Research and Development Center.

REFERENCES

Bates, B. C., and E. P. Campbell. 2001. A Markov chain Monte Carlo scheme for parameter estimation and inference in conceptual rainfall-runoff modeling. *Water Resources Research* 37(4):937–947.

Beven, K., and A. Binley. 1992. The future of distributed models: Model calibration and uncertainty prediction. *Hydrological Processes* 6(3):279–298.

Boyle, D. P., H. V. Gupta, and S. Sorooshian. 2000. Toward improved calibration of hydrologic models: Combining the strengths of manual and automatic methods. *Water Resources Research* 36(12):3663–3674.

Briggs, A. H., M. C. Weinstein, E. A. L. Fenwick, J. Karnon, M. J. Sculpher, and A. D. Paltiel. 2012. Model parameter estimation and uncertainty analysis: A report of the ISPOR-SMDM Modeling Good Research Practices Task Force Working Group–6. *Medical Decision Making* 32(5):722–732.

Davis, D., B. A. Faber, and J. R. Stedinger. 2008. USACE experience in implementing risk analysis for flood damage reduction projects. *Journal of Contemporary Water Research and Education* 140:3–14.

Doherty, J. 2004. *PEST: Model-independent parameter estimation, User Manual.* 5th ed. Brisbane, Queensland, Australia: Watermark Numerical Computing.

Doherty, J., and J. M. Johnston. 2003. Methodologies for calibration and predictive analysis of a watershed model. *Journal of American Water Resources Association* 39(2):251–265.

Doherty, J., and B. E. Skahill. 2006. An advanced regularization methodology for use in watershed model calibration. *Journal of Hydrology* 327:564–577.

Gallagher, M., and J. Doherty. 2007. Parameter estimation and uncertainty analysis for a watershed model. *Environmental Modeling & Software* 22:1000–1020.

Hydraulic Engineer Center (HEC). 2006. *Hydrologic modeling system HEC-HMS: User's manual*. Davis, CA: U.S. Army Corps of Engineers, Hydrologic Engineering Center.

Hill, M. C. 1998. Methods and guidelines for effective model calibration. *U.S. Geological Survey Water-Resources Investigations Report 98-4005*. Denver, CO: U.S. Geological Survey.

Kuczera, G. 1983. Improved parameter inference in catchment models. 1. Evaluating parameter uncertainty. *Water Resources Research* 19(5):1151–1172.

Lee, P. M. 2012. *Bayesian statistics: An introduction.* 4th ed. Hoboken, NJ: Wiley.

Madsen, H. 2000. Automatic calibration of a conceptual rainfall-runoff model using multiple objectives. *Journal of Hydrology* 235:276–288.

Mandel, J. 1984. The statistical analysis of experimental data. New York: Dover Publications.

Matott, L. S., J. E. Babendreier, and S. T. Purucker. 2009. Evaluating integrated environmental models: A survey of concepts and tools. *Water Resources Research* 45:W06421.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. Equation of state calculations by fast computing machines. *Journal of Chemical Physics* 21:1087–1092.

Mishra, S., N. E. Deeds, and B. S. RamaRao. 2003 Application of classification trees in the sensitivity analysis of probabilistic model results. *Reliability Engineering & System Safety* 73:123–129.

Mishra, S. 2009. Uncertainty and sensitivity analysis techniques for hydrologic modeling. *Journal of Hydroinformatics* 11(3–4):282–296.

Robert, C. P., and G. Casella. 2004. *Monte Carlo statistical methods*. 2nd ed. New York: Springer.

Ross, J. L., M. M. Ozbek, and G. F. Pinder. 2009. Aleatoric and epistemic uncertainty in groundwater flow and transport simulation. *Water Resources Research* 45:W00B15.

Servat, E., and A. Dezetier. 1991. Selection of calibration objective functions in the context of rainfall-runoff modelling in a Sudanese savannah area. *Journal of Hydrological Sciences* 36(4):307–330.

Scharffenberg, W., and M. L. Kavvas. 2011. Uncertainty in flood wave routing in a lateral-inflow-dominated stream. *ASCE Journal of Hydrologic Engineering* 16(2):165–175.

Skahill, B. E. 2006. *Potential improvements for HEC-HMS automated parameter estimation*. ERDC/CHL TR-06-13. Vicksburg, MS: U.S. Army Engineer Research and Development Center.

Skahill, B. E. 2013a. *Practice driven and state-of-the-art methods to quantify hydrologic model uncertainty*. ERDC/CHL CHETN-IV-87. Vicksburg, MS: U.S. Army Engineer Research and Development Center.

Skahill, B. E. 2013b. *Initial development and testing of a state-of-the-art method to quantify hydrologic model uncertainty*. ERDC/CHL CHETN-VIII-7. Vicksburg, MS: U.S. Army Engineer Research and Development Center.

Skahill, B. E., J. S. Baggett, S. Frankenstein, and C. W. Downer. 2009. Efficient Levenberg-Marquardt method based model independent calibration. *Environmental Modelling & Software* (24):517–529.

Skahill, B. E., and J. S. Baggett. 2012. *More efficient Bayesian-based optimization and uncertainty assessment of hydrologic model parameters*. ERDC/CHL TR-12-4. Vicksburg, MS: U.S. Army Engineer Research and Development Center.

Srivastava, A. 2008. Generalized event tree algorithm and software for dam safety risk analysis. M.S. Thesis, Utah State University, Logan, UT.

Tonkin, M. J., and J. E. Doherty. 2009. Calibration-constrained Monte Carlo analysis of highly-parameterized models using subspace techniques. *Water Resources Research* 45:W00B10.

Tonkin, M. J., J. E. Doherty, and C. Moore. 2007. Efficient nonlinear predictive error variance for highly parameterized models. *Water Resources Research* 43:W07429.

U.S. Army Corps of Engineers (USACE). 1996a. *Risk based analysis for evaluation of hydrology/hydraulics, geotechnical stability, and economics in flood damage reduction studies*. ER 1105-2-101. Washington, DC: U.S. Army Corps of Engineers. Revised January 2006.

———. 1996b. *Guidance on levee certification for the national flood insurance program*. Policy Letter. Washington, DC: U.S. Army Corps of Engineers.

———. 1996c. *Risk-based analysis of flood damage reduction studies*. EM 1110-2-1619. Washington, DC: U.S. Army Corps of Engineers.

Vanrolleghem, P. A., ed. 2010. *Modelling aspects of water framework directive implementation,* Vol. 1. New York: IWA Publishing.

Vrugt, J. A., C. J. F. ter Braak, H. V. Gupta, and B. A. Robinson. 2008. Equifinality of formal (DREAM) and informal (GLUE) Bayesian approaches in hydrologic modeling? In *Stochastic Environmental Research and Risk Assessment*. Berlin: Springer-Verlag.